



# UNIVERSIDAD COMPLUTENSE MADRID

Proyecto de Innovación

Convocatoria 2018/2019

**Nº de proyecto:**

166

**Título del proyecto:**

Tutorial interactivo de ejemplos básicos y ejercicios de Inferencia Estadística No-paramétrica mediante software libre: implementación mediante R, R-studio y Swirl

**Nombre del responsable del proyecto:**

Nirian Martín Apaolaza

**Centros:**

Facultad de Comercio y Turismo, Facultad de Ciencias Matemáticas

**Departamentos:**

Departamento de Economía Financiera y Actuarial y Estadística, Departamento de Estadística e Investigación Operativa

## Contenido

1. Objetivos propuestos en la presentación del proyecto .....	3
2. Objetivos alcanzados .....	4
3. Metodología empleada en el proyecto .....	6
4. Recursos humanos .....	7
5. Desarrollo de las actividades .....	7
5.1 Profesor: Creación de una lección .....	7
5.2 Alumno: Realización de una lección .....	8
6 Anexo .....	9
6.1 Profesor: Creación de una lección .....	9
6.2 Alumno: Realización de una lección .....	10

## 1. Objetivos propuestos en la presentación del proyecto

El objetivo principal del proyecto es construir un nuevo material docente para las asignaturas con contenido de Inferencia Estadística No-paramétrica de las Facultades de Ciencias Matemáticas y de Ciencias Económicas y Empresariales; como, por ejemplo, la asignatura de Inferencia Estadística que se imparte en la Facultad de Ciencias Matemáticas, en el cuarto curso del del doble grado de Economía y Matemáticas-Estadística, así como en el tercer curso del grado de Matemáticas-Estadística. Este es una ayuda para los estudiantes a la hora de manejar Software estadístico, en particular R, y las herramientas R-Studio y “Swirl” R.

Siguiendo este objetivo principal, como objetivos específicos del presente Proyecto, se han planteado, y resuelto:

- 1) Familiarizar al alumno y al profesor en el manejo del paquete estadístico R y fomentar su utilización.
- 2) Introducir una herramienta de motivación “innovadora” que facilite la comprensión de la abstracción en la parte teórica de la asignatura de Inferencia Estadística.
- 3) Crear ejemplos con salidas del programa R, mediante “Swirl”.
- 4) Crear material práctico de Laboratorio en la asignatura de Inferencia Estadística del doble grado de Matemáticas-Estadística y Economía, así como el grado de Matemáticas-Estadística, ya que hasta la actualidad sólo existe material basado en ejercicios de pizarra.

El uso del Software R ha ido popularizándose en los últimos años debido a su facilidad de sintaxis, su clara orientación estadística y, sobre todo, al hecho de tratarse de un Software libre, que continuamente sigue actualizándose con la inclusión de numerosos paquetes y librerías. Hoy en día, es indispensable, a nivel académico, ofrecer a los alumnos una formación práctica que complemente a la clásica formación puramente teórica.

“Swirl” es un paquete del programa estadístico R que transforma la consola de comandos de R en un entorno de aprendizaje interactivo, de modo que se puede aprender R usando R. Los usuarios reciben una interacción inmediata guiándoles en temas a un ritmo individual para seguir los contenidos de Estadística y programación en R. La mayor parte del contenido original de los cursos “Swirl” han sido desarrollados centrándose en estudiantes principiantes, pero existe también contenido más avanzado y continuará añadiéndose aún más material didáctico en el futuro. Los cursos van cargándose y actualizándose constantemente en la red de cursos. Hasta ahora, sin embargo, no conocemos la existencia de un curso de Inferencia Estadística No-Paramétrica en “Swirl”. Es por tanto que se considera necesario trabajar en este tema en particular.

## 2. Objetivos alcanzados

Pese a la facilidad de uso de “Swirl”-R, son necesarios ciertos conocimientos básicos del paquete estadístico R. Sin embargo, la mayoría de los alumnos de la Facultad de Ciencias Matemáticas carecen de formación en el uso de Software Estadístico. En particular, no hay ninguna asignatura obligatoria en el grado de Matemáticas cuyo plan de estudios prevea su aprendizaje; y en los grados de Matemáticas-Estadística y doble grado de Economía y Matemáticas-Estadística, el uso de Software estadístico (por ejemplo, R) no se da hasta tercer y cuarto curso, respectivamente.

Tras observar esta carencia, se consideró que lo primero de todo era importante ofrecer a los alumnos un curso gratuito en el que se diera una Introducción a R. Este se realizó en el ciclo de cursos de Compumates (<https://matematicas.ucm.es/compumates>) y tuvo una duración de 10 horas. Uno de los profesores encargados de este curso fue Elena Castilla, miembro del Proyecto (El otro profesor, Pedro Chocano, se adhiere al equipo de trabajo en el nuevo proyecto concedido, véase Sección 4). Con la realización de este curso se pretendía:

- 1) Dar una formación básica de R a los alumnos de la Facultad de Ciencias Matemáticas (uno de los objetivos planteados en el proyecto), y
- 2) Detectar necesidades particulares de los alumnos de Ciencias Matemáticas a la hora de aprender R, para intentar que se vieran reflejadas en el curso de Swirl R a desarrollar.

Ambos objetivos se vieron satisfechos a la vista de las encuestas realizadas al terminar el curso. Así, 14 de los 15 estudiantes que realizaron la encuesta consideraron el curso “Útil” o “Muy útil”. A la pregunta “¿Se lo recomendarías a un amig@?”, 9 de los 15 estudiantes respondieron “Por supuesto” (Ver Figura 1). Estas encuestas reflejan también, la alta predisposición de los alumnos de la Facultad a aprender R.

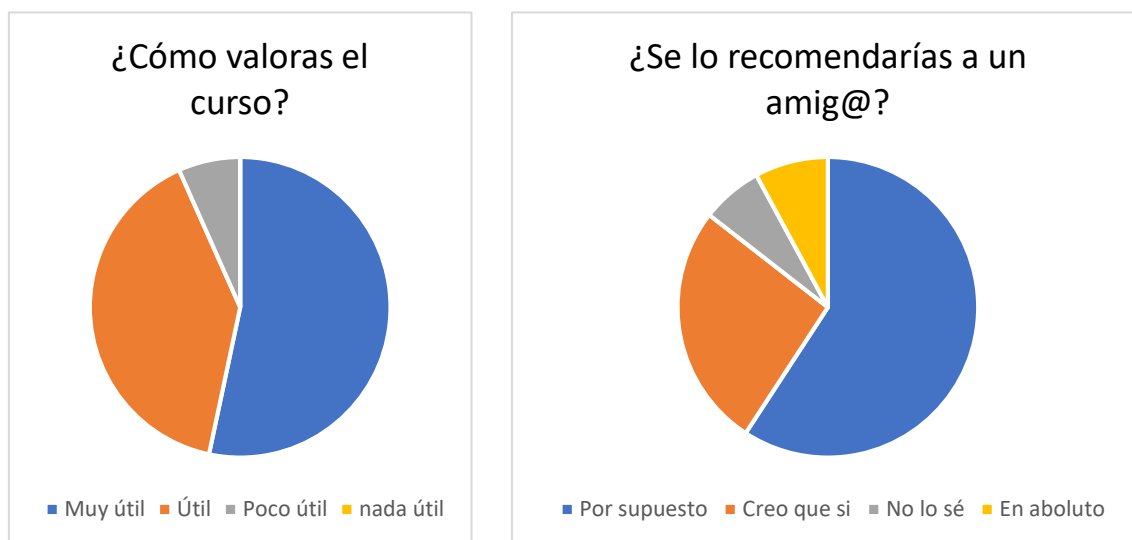


Figura 1: Algunos resultados de las encuestas realizadas en el Curso de Introducción a R (Facultad de Ciencias Matemáticas, Compumates 2018/2019). Se observa una gran satisfacción entre los alumnos del curso ofertado y cómo hay una gran predisposición a aprender R entre los alumnos de la Facultad.

Por otra parte, y siguiendo el objetivo 2) de detectar aquellas necesidades de los alumnos a la hora de aprender R, vimos como hubo muchos comentarios en las encuestas finales solicitando más material para el aprendizaje de éste. Algunos de los comentarios recibidos fueron los siguientes (Figura 2)

*"me gustaría disponer de más ejemplos"*  
*"me gustaría disponer de ejercicios resueltos"*  
*"me gustaría que se explicaran las soluciones de los ejercicios más despacio"*

Figura 2: Algunos comentarios en las encuestas realizadas en el Curso de Introducción a R (Facultad de Ciencias Matemáticas, Compumates 2018/2019). Los alumnos requieren de ejemplos y material explicativo para comenzar a aprender R.

Es decir: los alumnos quieren ejemplos, ejemplos resueltos y disponer de la explicación de las soluciones "a su ritmo". Con esos objetivos se planteó el desarrollo del material de "Swirl" R desarrollado, justificando los puntos 2) 3) y 4) de los objetivos específicos planteados.

Comentar aquí, que, durante el curso de Introducción a R, se preguntó a los alumnos si conocían "Swirl" R. Dos alumnos contestaron afirmativamente y dijeron y que les parecía una herramienta útil y dinámica.

Se ha desarrollado una librería en "Swirl", llamada **Inferencia no paramétrica**, que, hasta el momento de la presentación de la memoria, consta de las siguientes lecciones:

- **Bondad de Ajuste:** estudio de las pruebas Chi-Cuadrado, Kolgomorov-Smirnov y Lilliefors para contrastar si una muestra viene de una distribución en concreto.
- **Contrastes de Localización - Test de Wilcoxon:** muestras no pareadas, aplicación del Test de Wilcoxon a muestras de diferentes tamaños y con diferentes hipótesis alternativas.

El material docente elaborado se ha realizado en castellano. En las secciones 3, 5 y 6, se detallará con más precisión el trabajo realizado en esta librería. Si bien se pretende seguir ampliando el material, entendemos que el material está listo para su uso en cualquiera de las asignaturas con contenido de Estadística No-Paramétrica, y en concreto para la asignatura Inferencia Estadística

(<https://matematicas.ucm.es/estudios/2019-20/grado-matematicasyestadistica-plan-800644>). Así mismo, se prevé su uso también en la nueva edición de Compumates prevista para el curso 2019/2020. Se prevé subir los resultados a la plataforma de Swirl cuando se tengan más lecciones completadas.

Así, los objetivos logrados se podrían resumir en los siguientes puntos

- Familiarizar al alumno en el manejo de R fomentando su utilización.
- Crear material práctico e innovador que facilite el aprendizaje de R, creando ejemplos con salidas del programa R, mediante "Swirl", y que se pueda enseñar en la asignatura de Inferencia Estadística, así como en otras asignaturas con contenido en Inferencia No-paramétrica.

### 3. Metodología empleada en el proyecto

Para alcanzar los objetivos descritos en las Secciones 1 y 2 se llevaron a cabo las siguientes actuaciones:

- 1) **Primera Fase:** Introducción a R y detección de las necesidades específicas de aprendizaje en los alumnos. Para eso se llevó a cabo el curso de Introducción a R descrito previamente.
- 2) **Segunda Fase:** Diseño del curso de “Swirl” a crear. Se tuvieron en cuenta las conclusiones obtenidas en la primera fase, y se tuvo que tomar decisiones como:
  - idioma del curso: se decidió que castellano porque la docencia en las asignaturas de Estadística de la UCM es principalmente en este idioma.
  - contenidos a introducir: de momento se tratan los temas de Bondad de Ajuste y Modelos de Localización
  - dificultad y estructura de los ejemplos: se concluyó que no era necesario para el aprendizaje que los ejemplos tuvieran una dificultad excesivamente elevada, pero si era necesario que fuesen muy ilustrativos.
- 3) **Tercera Fase:** Programación del Curso en “Swirl”. La propia red de cursos dispone de ciertas guías a la hora de implementar nuevos cursos (<https://swirlstats.com/instructors.html>)

Queda pendiente la **Cuarta Fase**, que sería la implementación del curso programado en asignaturas de los diferentes grados donde los miembros del proyecto dan docencia. Se pretende comenzar a hacerlo al comienzo del curso 2019/2020, en el primer cuatrimestre

Las fases citadas se han desarrollado mediante reuniones de los profesores participantes, trabajo personal y consulta de otros cursos ya implementados en “Swirl”.

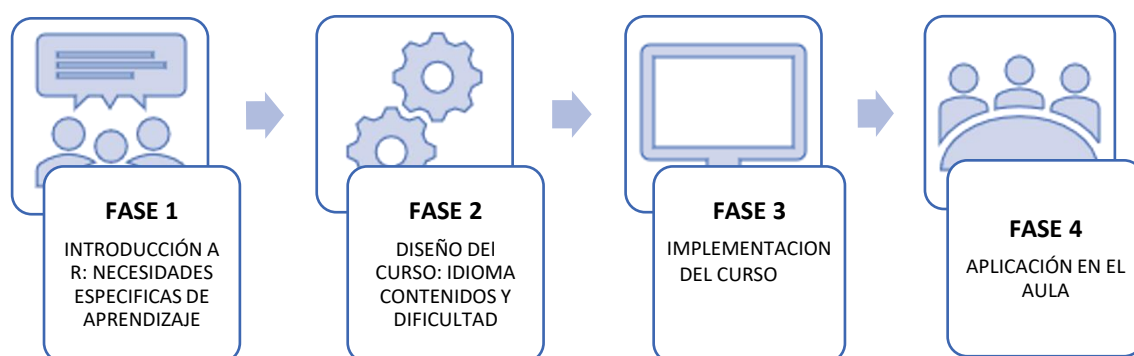


Figura 3: Esquema representativo de las fases seguidas en la realización del proyecto.

## 4. Recursos humanos

El presente proyecto, propuesto y coordinado por la Facultad de Comercio y Turismo, se trata de un proyecto interfacultativo e interdepartamental. En particular los centros y departamentos implicados son

- Facultad de Ciencias Matemáticas (Departamento de Estadística e Investigación Operativa);
- Facultad de Ciencias Económicas y Empresariales (Departamento de Economía Financiera y Actuarial y Estadística).

Los miembros de trabajo del proyecto pertenecen a las citadas facultades

- 1) **Nirian Martín Apaolaza**: Profesora Titular del Departamento de Economía Financiera y Actuarial y Estadística (responsable del proyecto).
- 2) **Elena María Castilla González**: Contratada predoctoral FPU del Departamento de Estadística e Investigación Operativa.
- 3) **Pedro Miranda Menéndez**: Profesor Titular del Departamento de Estadística e Investigación Operativa.
- 4) **Leandro Pardo Llorente**: Profesor Catedrático del Departamento de Estadística e Investigación Operativa.

Los miembros PDI del proyecto tienen una amplia experiencia en la docencia de asignaturas con contenidos de Inferencia Estadística y, en particular, de Inferencia Estadística No-paramétrica en los grados de Matemáticas, Ingeniería Matemática, Matemáticas y Estadística y doble grado de Economía y Matemáticas y Estadística, todos ellos en la Facultad de Ciencias Matemáticas; así como en el grado de Comercio y Turismo, en el cuál Nirian Martín, profesora responsable de este proyecto, imparte docencia como profesora Titular. Por otra parte, Leandro Pardo, es en la actualidad el coordinador y profesor encargado de impartir la asignatura de Inferencia Estadística en la Facultad de Matemáticas.

Dada los buenos resultados obtenidos, los miembros de este equipo de trabajo (junto con dos colaboradores más de la Facultad de Ciencias Matemáticas), han solicitado otro Proyecto de Innovación Docente para el curso 2019/2020. Este Proyecto, que ha sido aceptado, pretende la creación de manuales prácticos para el uso del Software Estadístico SAS, que complementará al material creado para R.

## 5. Desarrollo de las actividades

En esta sección se pretende explicar con más detalle como se ha desarrollado la librería **Inferencia no Paramétrica** en “Swirl” R

### 5.1 Profesor: Creación de una lección

Para crear material docente en “Swirl” hay que seguir los siguientes pasos

- 1) **Instalar R**: “Swirl” es un entorno pensado para enseñar a programar en R, por lo que será necesario tener R para instalarlo. En este sentido, los miembros del

proyecto hemos usado R para fines docentes e investigadores, por lo que no era necesario.

- 2) **Instalar R-Studio (recomendado):** El entorno R-Studio es siempre recomendable cuando se trabaja con R. Su interfaz permite ver los gráficos a la vez que se ejecuta código y, en este caso, a la vez que se realiza el curso. En este sentido cabe destacar, que en el curso de Introducción a R se instaba desde el principio a los alumnos a trabajar con R-Studio. En las imágenes mostradas en la Sección 6 podemos observar su apariencia.
- 3) **Instalar Swirl y Swirlify:** Esto nos permitirá crear nuevas lecciones.
- 4) **Crear un nuevo curso:** Este nuevo curso tendrá que ser guardado en la librería de Swirl en la carpeta de R del ordenador con el que se vaya a trabajar. La estructura básica de cada lección en el curso (la cual se guardará en una subcarpeta del curso) es la siguiente:
  - a. Fichero *lesson.yaml*: documento de texto que contiene toda la estructura de la lección. Hay diferentes clases de comandos en este fichero:
    - i. *text*: el más básico de todos, ejecuta un texto.
    - ii. *figure*: para ejecutar una figura, llamará al fichero de código de la figura para que se ejecute.
    - iii. *mult\_question*: se le pregunta al alumno una pregunta y tiene diferentes opciones para elegir.
    - iv. *cmd\_question*: se le pide al alumno que ejecute un comando.
  - b. Fichero *initLesson.R*: este fichero contendrá los comandos que se ejecutarán internamente nada más comenzar la lección.
  - c. Otros ficheros: se irán ejecutando a medida que avance la lección. Por ejemplo, dibujo de gráficos que ilustren los ejemplos y ejercicios.

En las Figuras 4 y 5 de la Sección 6.1 se muestra la estructura de la lección Bondad de Ajuste.

## 5.2 Alumno: Realización de una lección

Para hacer uso del material docente en “Swirl” hay que seguir los siguientes pasos

- 1) **Instalar R, R-Studio (recomendado) y Swirl:** es probable que los dos primeros estén instalados si el alumno ya tiene un conocimiento básico de R.
- 2) **Instalar el curso de Inferencia no Paramétrica:** en la librería de Swirl en la carpeta de R del ordenador en uso. Es tan sencillo como copiar la carpeta.
- 3) **Abrir Swirl y seleccionar idioma:** si queremos que las indicaciones del entorno Swirl sean en castellano (como el curso) hay que especificarlo con la sentencia `select_language(language = "spanish")`. Para empezar a interactuar con Swirl usaremos la sentencia `swirl()`.
- 4) **Empezar a usar swirl:** seleccionar la lección que se quiera cursar (no necesariamente la implementada en este proyecto) y seguir las instrucciones.

En las Figuras de 6 a 15 de la Sección 6.2 se muestra la realización de la lección Bondad de Ajuste.



## 6 Anexo

### 6.1 Profesor: Creación de una lección

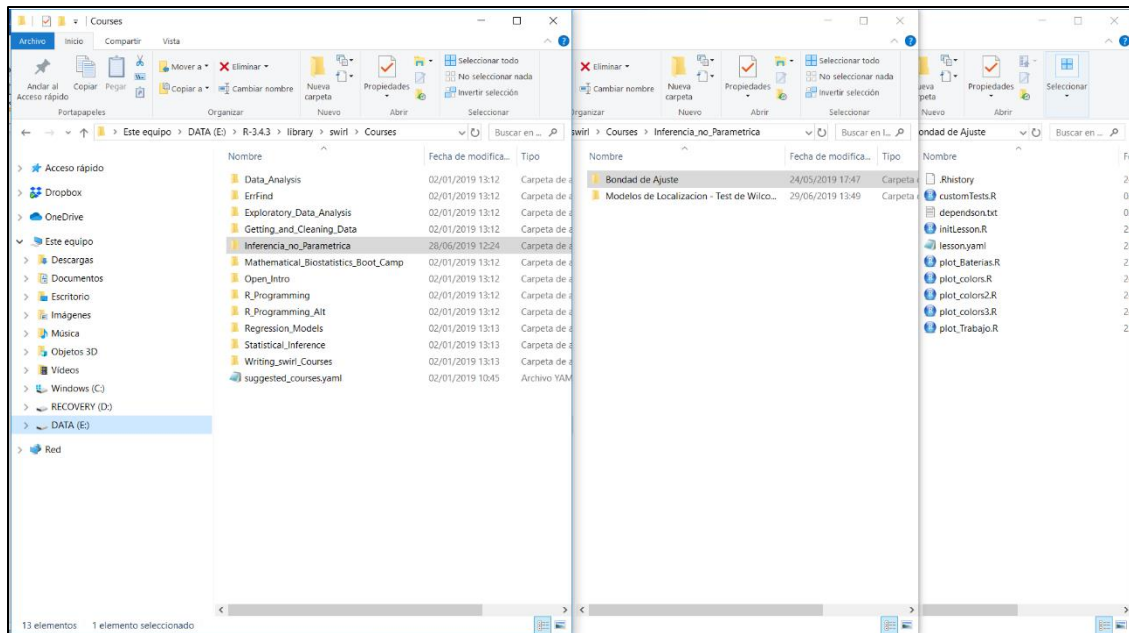


Figura 4: Carpeta del curso de Inferencia no Paramétrica, con las lecciones de Bondad de Ajuste y Contrastes de Localización-Test de Wilcoxon. Archivos de la Lección Bondad de Ajuste

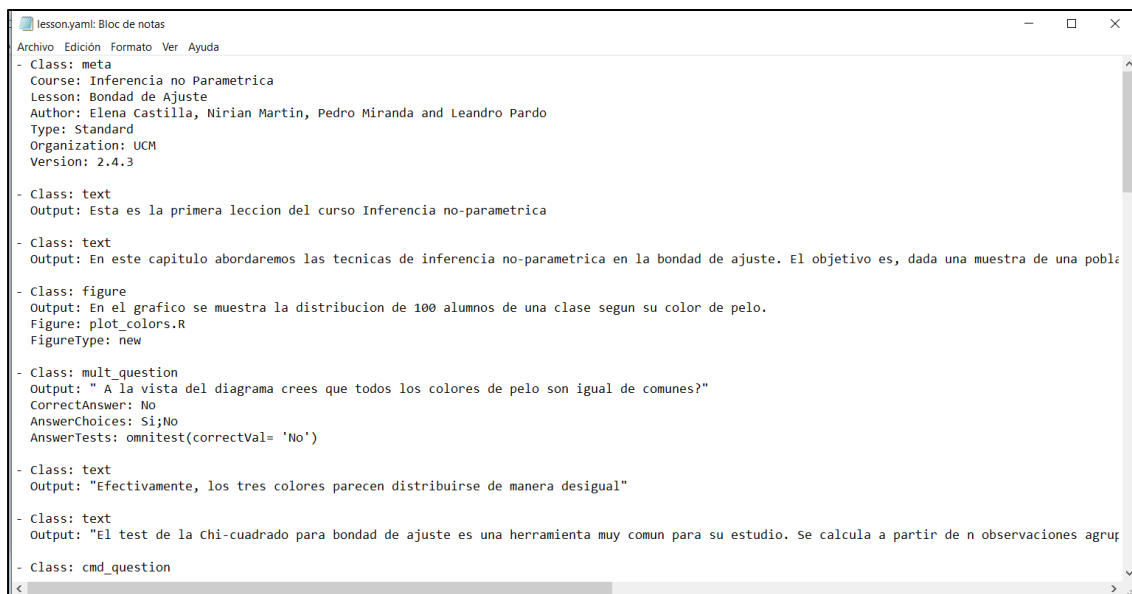


Figura 5: Curso de Bondad de Ajuste: fichero lesson.yaml.

## 6.2 Alumno: Realización de una lección

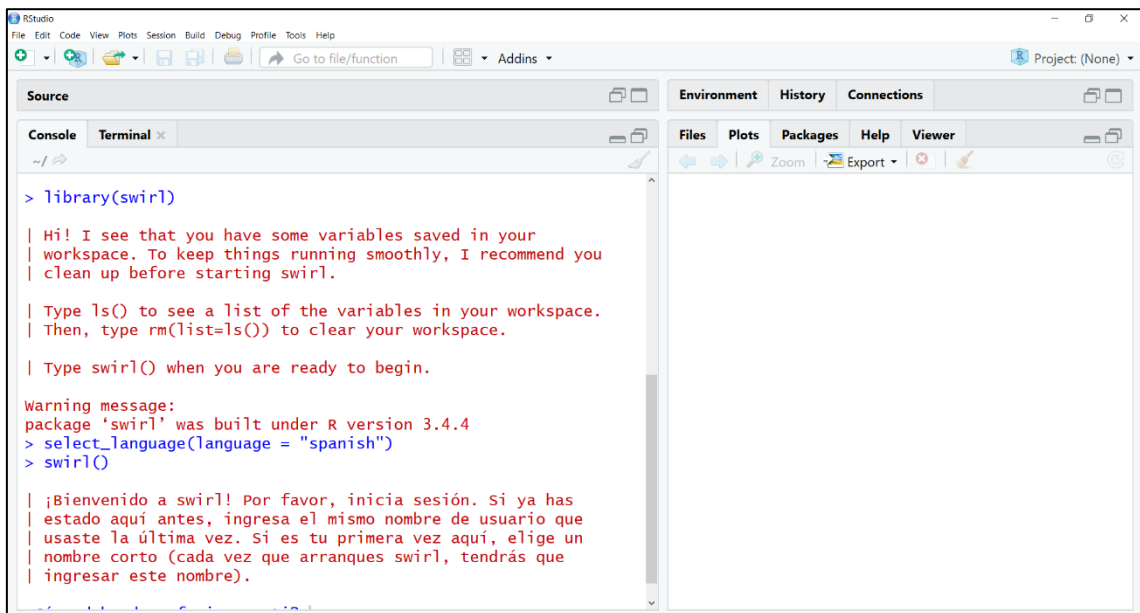


Figura 6: Inicialización de Swirl, selección de idioma

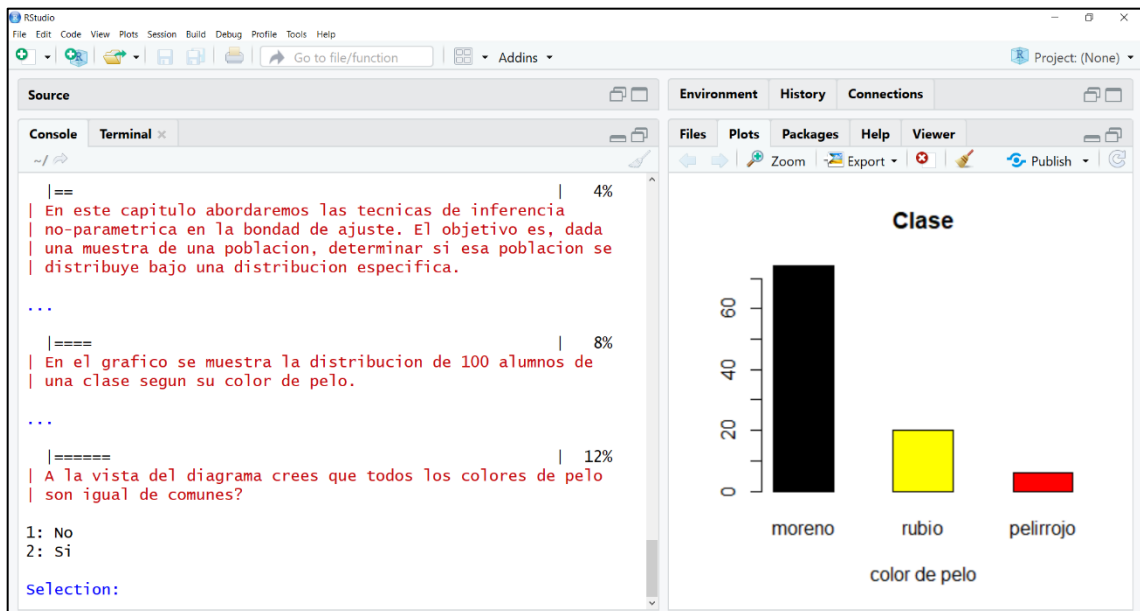


Figura 7: Lección de Bondad de Ajuste, primer ejemplo, colores de pelo en Clase

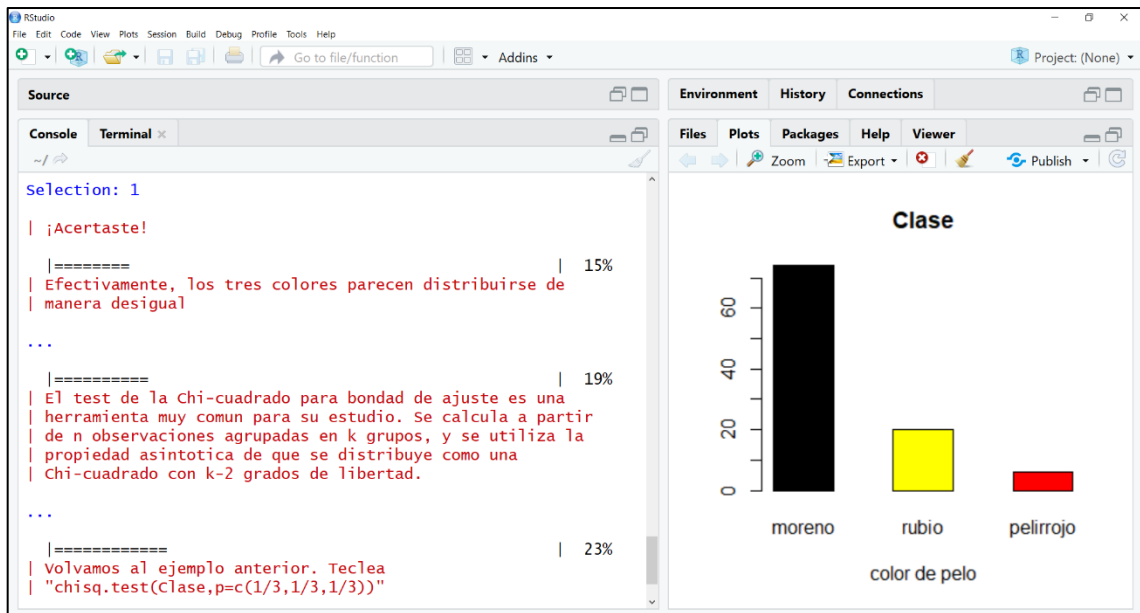


Figura 8: Lección de Bondad de Ajuste, primer ejemplo, colores de pelo en Clase, test de la Chi Cuadrado

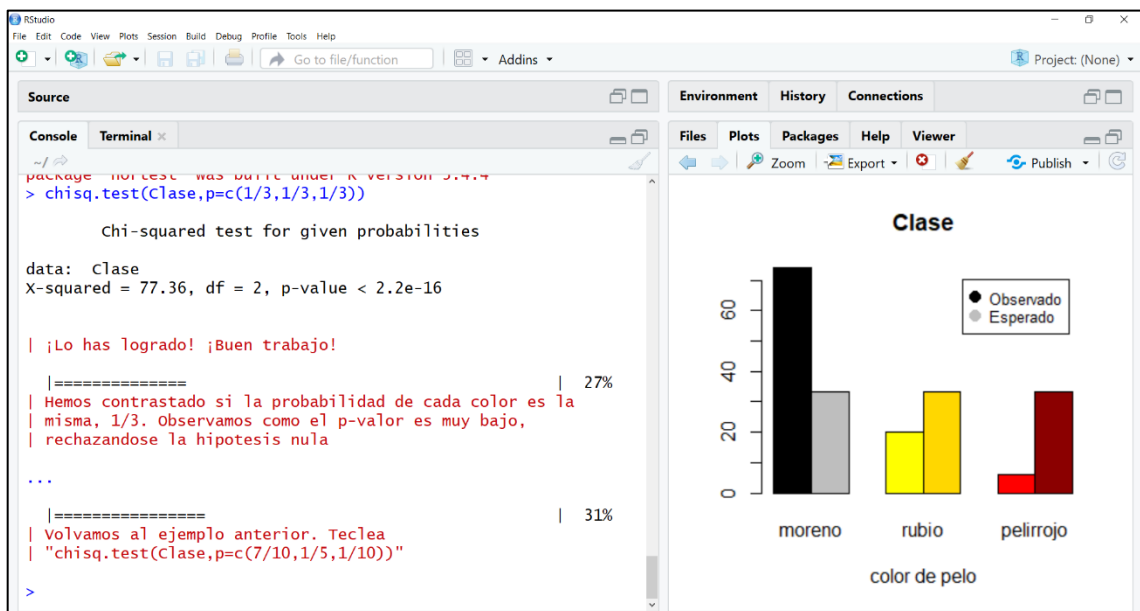


Figura 9: Lección de Bondad de Ajuste, primer ejemplo, colores de pelo en Clase, test de la Chi Cuadrado II

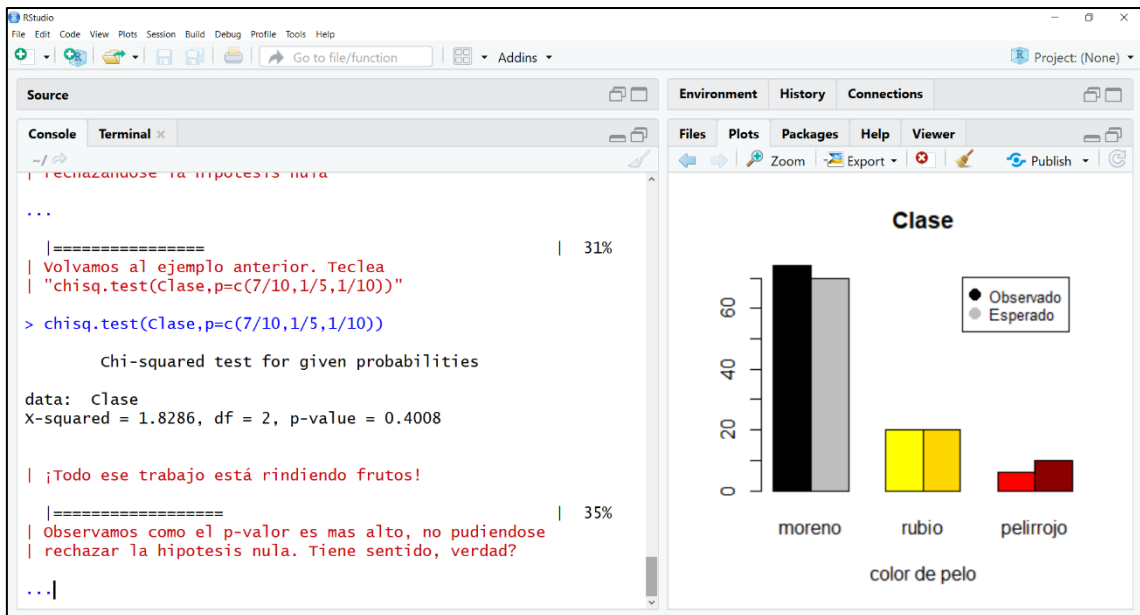


Figura 10 Lección de Bondad de Ajuste, primer ejemplo, colores de pelo en Clase, test de la Chi Cuadrado III

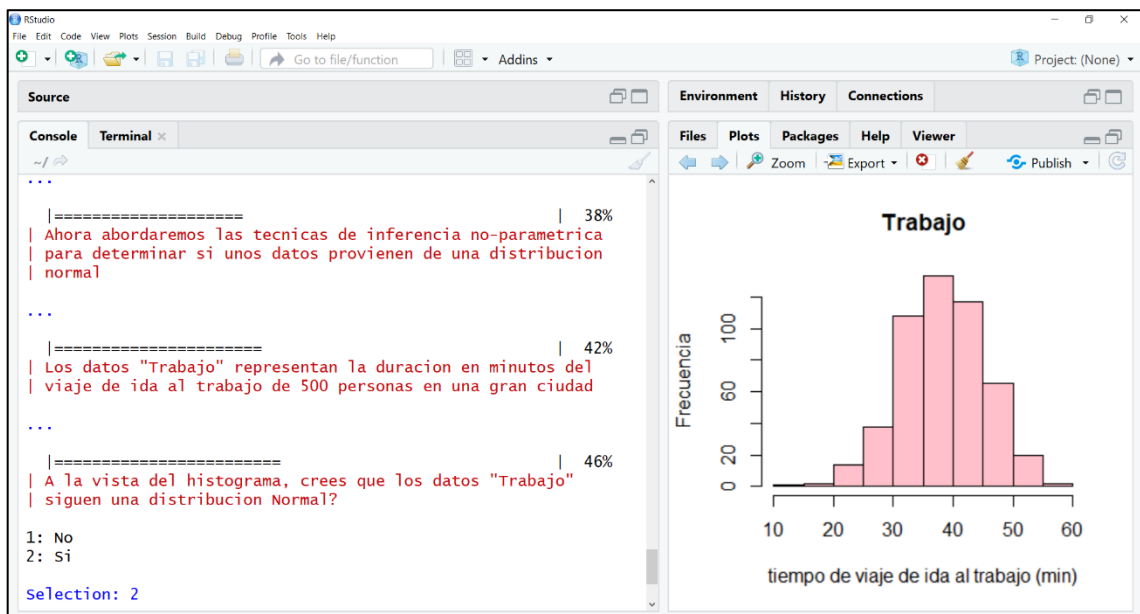


Figura 11 Lección de Bondad de Ajuste, segundo ejemplo, tiempo de viaje al trabajo

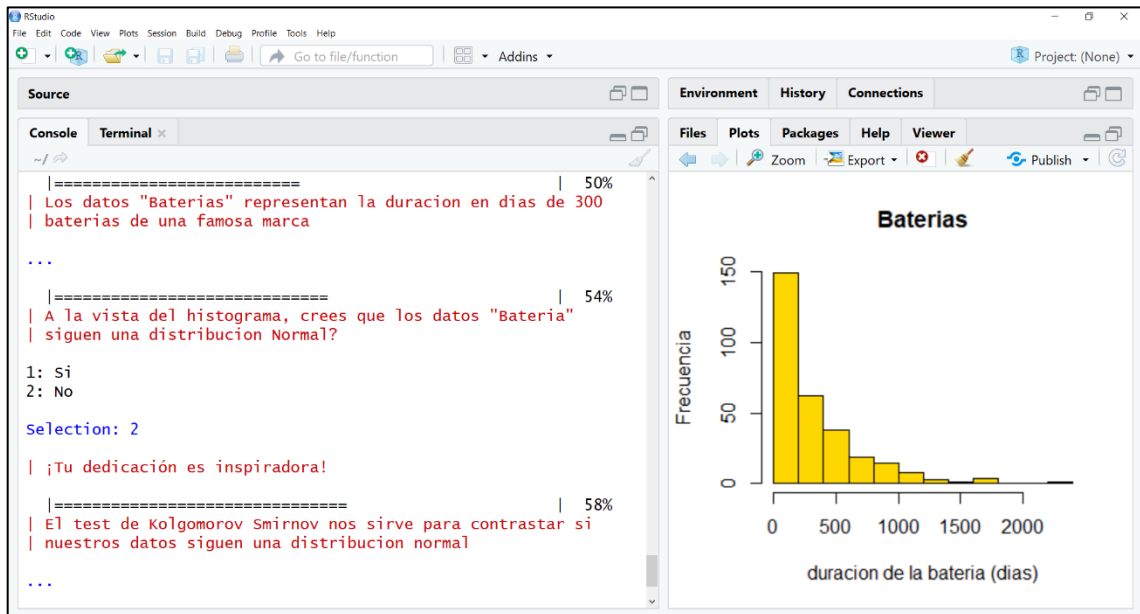


Figura 12 Lección de Bondad de Ajuste, segundo ejemplo, tiempo de duración de baterias

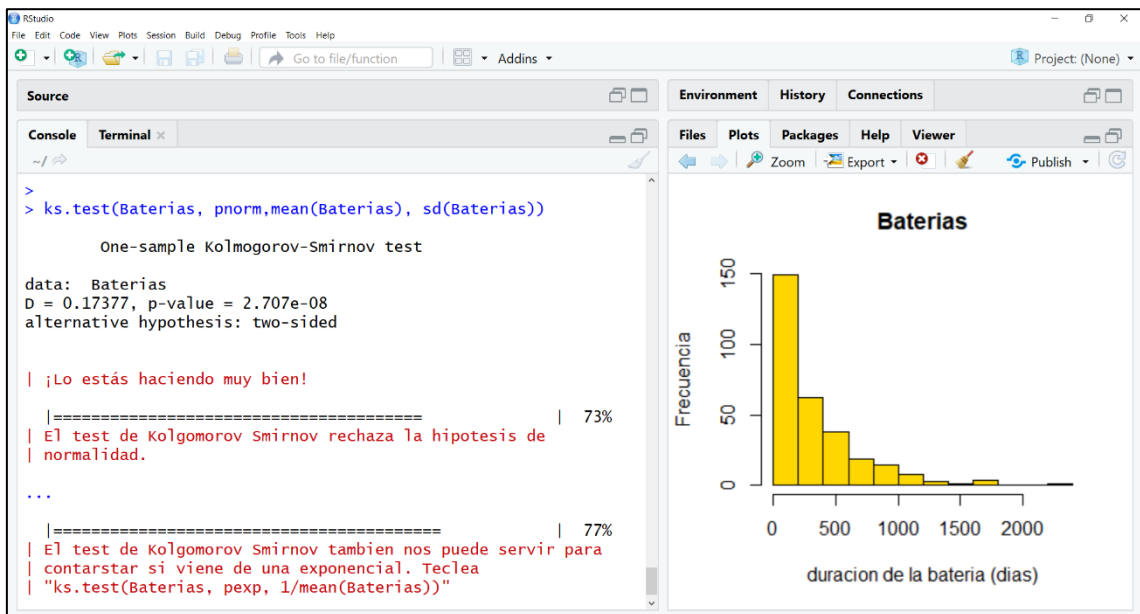


Figura 13 Lección de Bondad de Ajuste (continuación)

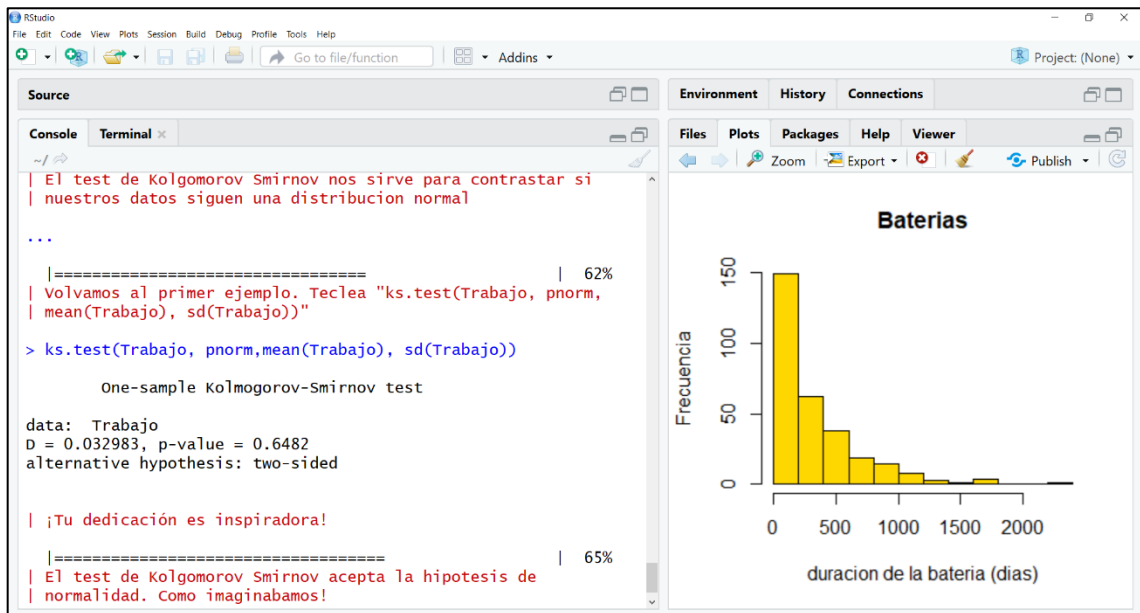


Figura 14 Lección de Bondad de Ajuste (continuación)

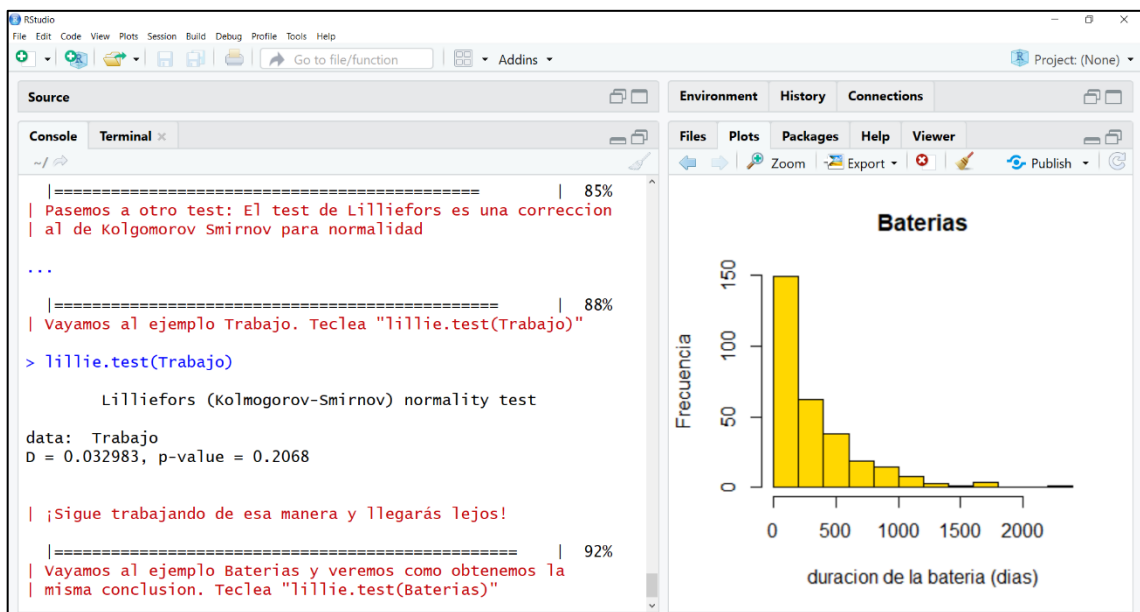


Figura 15 Lección de Bondad de Ajuste (continuación)